DOCUMENT RESUME

ED 047 020	24	TM 000 408

AUTHOR	Elashoff, Janet Dixon; Elashoff, Robert M.
TITLE	A Model for Quadratic Outliers in Linear Regression.
INSTITUTION	Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	RDM-69
BUREAU NO	BR-5-0252
PUB DATE	Dec 70
CONTRACT	OEC-6-10-078
NOTE	55p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Correlation, *Mathematical Models, *Multiple Regression Analysis, *Research Methodology, *Statistical Analysis, Statistical Data
IDENTIFIERS	*Outliers

ABSTRACT

This paper introduces a model for describing outliers (observations which are extreme in some sense or violate the apparent pattern of other observations) in linear regression which can be viewed as a mixture of a quadratic and a linear regression. The maximum likelihood estimators of the parameters in the model are derived and their asymptotic properties discussed. Small sample behavior of the model and robustness to inaccurate specification of the mixing parameter were investigated using Monte Carlo techniques. The asymptotic properties provide reasonable indications of behavior for n as small as 21 and the procedure appears quite robust to the inaccurate specification of the mixing parameter. Building models to describe outliers and estimating their parameters provides an interesting alternative to procedures of outlier detection followed by ordinary least squares procedures. (Author)

SC
RD
T

Stanford Center for Research and Development in Teaching

SCHOOL OF EDUCATION STANFORD UNIVERSITY

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

Research and Development Memorandum No. 69

A MODEL FOR QUADRATIC OUTLIERS IN LINEAR REGRESSION

Janet Dixon Elashoff
Stanford University

with the collaboration of

Robert M. Elashoff
University of California, San Francisco

School of Education
Stanford University
Stanford, California

December 1970

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the Office of Education. (Contract No. OE-6-10-078, Project No. 5-0252-0704.) Work on this study was also supported in part by National Institute of General Medical Science Grant No. GM 17182-02 SSS.

Introductory Statement

The central mission of the Stanford Center for Research and Development in Teaching is to contribute to the improvement of teaching in American schools. Given the urgency of the times, technological developments, and advances in knowledge from the behavioral sciences about teaching and learning, the Center works on the assumption that a fundamental reformulation of the future role of the teacher will take place. The Center's mission is to specify as clearly, and on as empirical a basis as possible, the direction of that reformulation, to help shape it, to fashion and validate programs for training and retraining teachers in accordance with it, and to develop and test materials and procedures for use in these new training programs.

The Center is at work in three interrelated problem areas: (a) Heuristic Teaching, which aims at promoting self-motivated and sustained inquiry in students, emphasizes affective as well as cognitive processes, and places a high premium upon the uniqueness of each pupil, teacher, and learning situation; (b) The Environment for Teaching, which aims at making schools more flexible so that pupils, teachers, and learning materials can be brought together in ways that take account of their many differences; and (c) Teaching Students from Low-Income Areas, which aims to determine whether more heuristically oriented teachers and more open kinds of schools can and should be developed to improve the education of those currently labled as the poor and the disadvantaged.

This paper grew out of the activities of the Center's Methodology Unit and represents a methodological development generated in answer to problems encountered in the reanalysis of the Rosenthal-Jacobson *Pygmalion in the Classroom* study. Such data analyses problems pose frequent difficulties in data gathered by Center projects.

Table of Contents

List of Tables

List of Figures

Abstract

This paper introduces a model for describing outliers in linear regression which can be viewed as a mixture of a quadratic and a linear regression. The maximum likelihood estimators of the parameters in the model are derived and their asymptotic properties discussed. Small sample behavior of the model and robustness to inaccurate specification of the mixing parameter were investigated using Monte Carlo techniques. The asymptotic properties provide reasonable indications of behavior for n as small as 21 and the procedure appears quite robust to the inaccurate specification of the mixing parameter. Building models to describe outliers and estimating their parameters provides an interesting alternative to procedures of outlier detection followed by ordinary least squares procedures.

INTRODUCTION

The standard linear regression model for fixed x's is given by

$$y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i \qquad i = 1, 2, \ldots, n \tag{1}$$

where

$$\mathrm{Cov}\ (\varepsilon_i \varepsilon_j) = 0 \qquad i \neq j \tag{2}$$

$$\varepsilon_i \sim N(0,\sigma^2) \tag{3}$$

Occasionally the data may contain observations inconsistent with the apparent pattern of the rest of the observations. Such aberrant observations or outliers could lead in extreme cases to rejection of (1) as the form of the regression relationship. Even if (1) is assumed, estimators of α and β by standard least squares procedures based on assumptions (2) and (3) may have unsatisfactory distributional properties such as large bias and large variance in the presence of outliers.

In this paper we formulate some models to describe outliers in regression problems, give a brief review of previous work in this area, and propose a particular model suggested by some real data. Then we derive the maximum likelihood estimators of the parameters in the model and their asymptotic properties. Monte Carlo investigations to determine the small sample properties of the maximum likelihood estimators and their robustness to inaccurate specification of the mixing parameter are reported. Large sample and small sample comparisons under our quadratic

outlier model of the maximum likelihood estimators and the ordinary least squares estimators for linear regression are discussed. Finally the model is applied to some data obtained in the Rosenthal-Jacobson teacher expectancy study.

Some Models for Outliers in Linear Regression

We begin by outlining some simple models for outliers in linear regression problems suggested by those proposed in the single sample case (see, e.g., Grubbs, 1969, or Dixon, 1962). Retaining assumptions (1) and (2), alternatives to (3) which generate outliers are models with skewed error distributions such as:

$$\varepsilon \sim (1 - \gamma)N(0,\sigma^2) + \gamma N(\lambda,\sigma^2) \tag{4}$$

$$\varepsilon \sim (1 - \gamma)N(0,\sigma^2) + \gamma N(\lambda(x),\sigma^2) \tag{5}$$

$$\varepsilon \sim (1 - \gamma(x))N(0,\sigma^2) + \gamma(x)N(\lambda,\sigma^2) \tag{6}$$

$$\varepsilon \sim (1 - \gamma(x))N(0,\sigma^2) + \gamma(x)N(\lambda(x),\sigma^2) \tag{7}$$

models like (4) or (5) in which it is known that $n-k$ of the ε observations are $N(0,\sigma^2)$ and that k of the observations are $N(\lambda,\sigma^2)$ or $N(\lambda(x),\sigma^2)$. (8)

With assumptions 1 and 2, error model (4) describes a process in which there is a constant probability that a y observation will be biased by an amount λ. In model (5) the probability is constant but the amount of bias depends on x. In model (6) the bias is constant but the probability

of a biased observation depends on x . Model (7) is a combination of (5) and (6).

We can propose analogous models with symmetric error distributions for the scale contaminated case:

$$\varepsilon \sim (1 - \gamma)N(0,\sigma^2) + \gamma N(0,\lambda^2\sigma^2) \tag{9}$$

$$\varepsilon \sim (1 - \gamma)N(0,\sigma^2) + \gamma N(0,\lambda^2(x)\sigma^2) \tag{10}$$

$$\varepsilon \sim (1 - \gamma(x))N(0,\sigma^2) + \gamma(x)N(0,\lambda^2\sigma^2) \tag{11}$$

$$\varepsilon \sim (1 - \gamma(x))N(0,\sigma^2) + \gamma(x)N(0,\lambda^2\sigma^2) \tag{12}$$

models like (9) or (10) in which it is known that $n-k$ of the ε observations are $N(0,\sigma^2)$ and k are $N(0,\lambda^2\sigma^2)$ or $N(0,\lambda^2(x)\sigma^2)$. (13)

models like (9) where $\lambda^2\sigma^2$ follows some distribution. (14)

Model (9) describes a process in which occasional y observations come from a population with a larger variance. In model (10) the variance of aberrant y observations depends on x . In model (11) the probability that a y observation has a larger variance depends on x . Model (12) is a combination of (10) and (11).

These models with $\lambda(x)$, $\gamma(x)$ suitably defined can describe a wide variety of cases.

Review of Literature

We define an outlier as an observation which is extreme in some sense or violates the apparent pattern of the other observations. Most of the statistical literature on outliers is concerned with two basic problems: detection of outliers and estimation of parameters in the presence of outliers.

There are several approaches to the detection problem when we have two variables. Let y and x denote the two variables and suppose at first that both y and x are random variables. For bivariate and multivariate models where x or y are distributed as in (8) or (13) with at most one outlier, a test statistic for outlier detection which maximizes the probability of making the correct decision has been discussed; see Ferguson (1961b), Karlin and Truax (1960). When more than one outlier is suspected there is little information on how to proceed. One technique is to apply the method described above repeatedly. Another is to have some prior information that particular observations are suspect and, then, possibly apply tests developed by Wilks (1963) that generalize Grubbs (1950). Still another alternative is to treat each variable separately and apply univariate single sample techniques.

When x is the independent variable and is measured without error and the regression of y on x is given by (1) where ϵ_i are distributed as in one of models (4)-(14), a number of suggestions for locating possible outliers have been made in the literature. One suggestion is to compute the maximum squared studentized residual and reject the observation corresponding to this residual if it is significantly large.

Clearly, this procedure has its difficulties; see Mickey et al. (1967). Another suggestion made by Mickey et al. (1967) is to find the single observation whose deletion causes the greatest reduction in the sum of squared residuals. Having found and deleted this observation, the procedure finds the next observation whose deletion reduces the sum of squared residuals as much as possible. No theory for the procedure is available. The procedure can be carried out on the computer by using a standard stepwise regression program such as BMD02R. The regression relation must have a known form (e.g., linear); but no distributional assumptions need to be made for x and the distribution of y may follow any of the models outlined above.

The problem of detecting outliers in the regression setup requires much more work. Little theoretical guidance for consumers of statistical regression analysis is available. A very interesting approach to outliers in calibration analysis is suggested by Youden (see Barnett, 1965).

A review of how to estimate α,β and a measure of their variability in the general case becomes a rather large problem. In our brief review we will restrict consideration to the model defined by (1), (2) and some choice of (4)-(13). So far the only work appears to be for symmetric error models such as those in (9)-(13).

The main lines of attack on the problem of choosing estimators for α,β are essentially generalizations of the approaches to the single sample problem. Anscombe's (1967) paper applies when the ϵ_i are a random sample from a t distribution or a distribution in some sense well-approximated by a t (an example of model (14)). Anscombe indicates

that minimization of the Huber metric may be used and, generally, will give estimates "close" to those obtained by his Bayesian approach using the t as the basic data distribution. If the ε_i are distributed as a scale contaminated compound normal distribution (model (9)), then the methods of Box and Tiao (1968) may be extended to derive estimators for α, β . Anscombe's (1960a) paper is useful when we want to test for skewness, kurtosis or heteroscedasticity. A few suggestions on estimation procedures for α, β based on ranks or signs have been investigated, see Mood (1950), Adichie (1967a), Sen (1968), Theil (1950).

Estimators for α, β may also be deduced by first screening the data for outliers by one of the techniques suggested in the section on detection and then estimating α, β by minimizing some metric. Not much is known about this approach except the paper by Anscombe and Barron (1966) for estimating the population mean from a single sample.

A QUADRATIC OUTLIER MODEL

Our interest in the problem of outliers in linear regression problems was kindled by two examples of data problems in which aberrant observations seemed to occur only on one side of the regression line and at one extreme of the x's (see Figures 5, 6, 7). Thus we were led to consideration of error models (4)-(7). Model (5) seemed to describe best our impression that outliers were increasingly far from the line for more extreme x and we were led to an examination of model (5) with a reasonable specification of $\lambda(x)$.

This paper, then, is concerned with estimation of the parameters in the quadratic outlier model (15). Since the adoption of such a model implies the occurrence of a similar pattern of outliers across several

sets of data and the model may generate many nonextreme "aberrant" observations, it seems more profitable to concentrate our efforts directly on parameter estimation rather than on any two-stage detection of outliers and parameter estimation procedures.

Quadratic outlier model:

$$y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i \qquad i = 1, 2, \ldots, n$$

$$\text{Cov}\ (\varepsilon_i, \varepsilon_j) = 0 \qquad i \neq j$$

$$\varepsilon \sim (1 - \gamma)N(0,\sigma^2) + \gamma N(\lambda(x),\sigma^2) \tag{15}$$

$$\lambda(x) = c\ (x - m)^2$$

m and γ known

x's fixed

We choose $\lambda(x) = c\ (x_i - m)^2$ with m known, as a simple function which describes our impression of the data. We assume that the general pattern of outliers, and thus m, is known. Model (15) describes a bias which increases rapidly for extreme x. By defining m as x_{min}, $\bar{x}$, x_{max} and forcing c to be positive or negative we can obtain the bias patterns shown in Figure 1.

The assumption of known γ is not so restrictive as might at first appear. The literature indicates that its accurate estimation is difficult and our own results indicate that incorrect specification is

not serious. The problem of estimating the parameters of a mixture of distributions has been around a long time. Pearson (1894) discussed estimates based on the method of moments. Rao (1952) reviewed this approach but pointed out that the estimate of the proportion of the mixture has a large variance and its estimation requires very large samples. Hill (1963), using some expansions of the information for the estimation of the mixing probability γ for two exponential distributions, shows that unless the mixed distributions are very well separated, extremely large samples are needed even for moderate precision when all other parameters are known. Larger samples are needed if the other parameters must be estimated as well. Box and Tiao (1968) exploring the estimation of θ in the mixture $(1 - \gamma)N(\theta,\sigma^2) + \gamma N(\theta,k^2\sigma^2)$ by Bayesian methods assuming k and γ known and then using various values of k and γ showed that the estimator of θ is not unduly sensitive to changes in k or γ in a reasonable range.

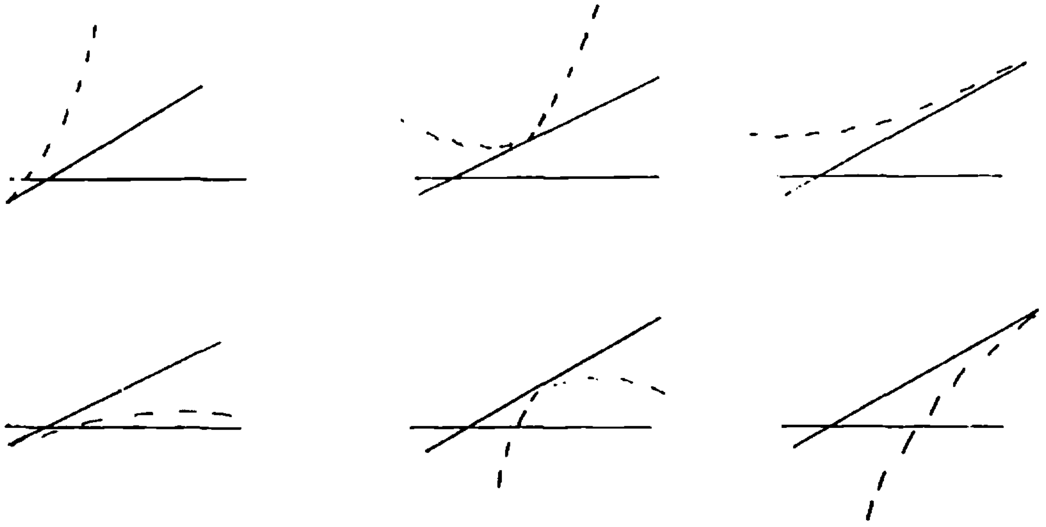

Figure 1: Bias patterns generated by model (15)

Maximum Likelihood Estimators

Assuming m specified, γ known, and the x's fixed, the maximum likelihood estimators of α, β, c, σ^2 are given by the following equations:

$$\hat{\alpha} = \bar{y} - \frac{\gamma\hat{c}}{n}\Sigma(x_i - m)^2 w_i \tag{16}$$

$$\hat{\beta} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} - \gamma\hat{c}\frac{\Sigma(x_i - \bar{x})(x_i - m)^2 w_i}{\Sigma(x_i - \bar{x})^2} \tag{17}$$

$$\hat{\sigma}^2 = \frac{\Sigma(y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x}))^2}{n} - \frac{\gamma\hat{c}^2}{n}\Sigma(x_i - m)^4 w_i \tag{18}$$

$$\hat{c} = \frac{\Sigma(x_i - m)^2 (y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x})) w_i}{\Sigma(x_i - m)^4 w_i} \tag{19}$$

where

$$w_i^{-1} = \gamma + (1-\gamma)e^{\frac{1}{2\hat{\sigma}^2}[-2\hat{c}(x_i - m)^2(y_i - \hat{\alpha} - \hat{\beta}(x_i-\bar{x})) + \hat{c}^2(x_i-m)^4]} \tag{20}$$

A Fortran IV computer program to obtain iterative solutions to these equations was written.

Asymptotic Properties of the Maximum Likelihood Estimators

Asymptotically the maximum likelihood estimators of α, β, c, σ^2 have a multivariate normal distribution. That is, for fixed x's in the interval (a, b) the estimators

$$\sqrt{n}\,(\hat{\alpha}_{ML}-\alpha)\ ,\quad \sqrt{n}\,(\hat{\beta}_{ML}-\beta)\ ,\quad \sqrt{n}\,(\hat{c}_{ML}-c)\ ,\quad \sqrt{n}\,(\hat{\sigma}^2_{ML}-\sigma^2)$$

have an asymptotic multivariate normal distribution with variance-covariance matrix $\underset{\sim}{V}$ given by $n\,\underset{\sim}{M}^{-1}$ where $\underset{\sim}{M}$ is the information matrix. Letting $\underset{\sim}{A} = \sigma^4\underset{\sim}{M}$, the terms in A are given by the following formulas:

$$a_{11} = n\sigma^2 - \gamma(1-\gamma)c^2\Sigma(x_i-m)^4 I_i$$

$$a_{12} = -\gamma(1-\gamma)c^2\ \Sigma(x_i-m)^4(x_i-\bar{x})I_i$$

$$a_{13} = \gamma\sigma^2\ \Sigma(x_i-m)^2 + \gamma(1-\gamma)c\Sigma(x_i-m)^4\ [J_i-c(x_i-m)^2 I_i]$$

$$a_{14} = \frac{\gamma(1-\gamma)}{\sigma^2}\,c^2\ \Sigma(x_i-m)^4\ (-J_i + \frac{c}{2}\,(x_i-m)^2 I_i)$$

$$a_{22} = \sigma^2\ \Sigma(x_i-\bar{x})^2 - \gamma(1-\gamma)c^2\Sigma(x_i-m)^4\ (x_i-\bar{x})^2 I_i$$

$$a_{23} = \sigma^2\gamma\Sigma(x_i-\bar{x})(x_i-m)^2 + \gamma(1-\gamma)c\Sigma(x_i-\bar{x})(x_i-m)^4[J_i-c(x_i-m)^2 I_i]$$

$$a_{24} = \frac{\gamma(1-\gamma)}{\sigma^2}\,c^2\Sigma(x_i-\bar{x})(x_i-m)^4\ (-J_i + \frac{c}{2}\,(x_i-m)^2 I_i)$$

$$a_{33} = \gamma\sigma^2\Sigma(x_i-m)^4 - \gamma(1-\gamma)\Sigma(x_i-m)^4\ (K_i-2c(x_i-m)^2\ J_i+c^2(x_i-m)^4 I_i)$$

$$a_{34} = \frac{\gamma(1-\gamma)}{\sigma^2}\,c\Sigma(x_i-m)^4\ (K_i - \frac{3}{2}c(x_i-m)^2\ J_i + \frac{c^2}{2}\ (x_i-m)^4 I_i)$$

$$a_{44} = \frac{n}{2} - \frac{\gamma(1-\gamma)}{\sigma^4}\,c^2\Sigma(x_i-m)^4\ [K_i - c(x_i-m)^2\ J_i + \frac{c^2}{4}\ (x_i-m)^4 I_i)$$

where

$$I_i = \frac{1}{\sqrt{2\pi}\sigma} \int e^{-z_i^2/2\sigma^2} f(z_i)\, dz_i$$

$$J_i = \frac{1}{\sqrt{2\pi}\sigma} \int z_i\, e^{-z_i^2/2\sigma^2} f(z_i)\, dz_i$$

$$K_i = \frac{1}{\sqrt{2\pi}\sigma} \int z_i^2\, e^{-z_i^2/2\sigma^2} f(z_i)\, dz_i$$

and

$$f(z_i) = [(1-\gamma)e^{\frac{-c(x_i-m)^2}{2\sigma^2}[2z_i - c(x_i-m)^2]} + \gamma]^{-1} .$$

To demonstrate the way in which the asymptotic variances vary with the parameters γ, f and to gain an idea of variances we might expect in small samples we evaluated the formulas for the asymptotic variances of $\hat{\alpha}$, $\hat{\beta}$, $\hat{c}$, $\hat{\sigma}^2$ at several values of n (these numbers are then taken from $\underset{\sim}{M}^{-1}$). As we shall see in later sections these asymptotic formulas for the variances may provide very good approximations to the actual variances for n's as small as 21.

Asymptotic formulas for the variances were computed for some illustrative cases. We set $\alpha = 0.0$, $\beta = 1.0$, $m = x_{min}$, and c positive. The x's were equally spaced from -1.0 to +1.0 with one y observation at each x . The asymptotic variances were evaluated for sample sizes, n , of 15, 21, 41 and γ of .10, .20, .30, .40. Values

of σ^2 were chosen so that $\frac{\sigma^2}{\sigma_y^2} = .50, .20$; that is, values of σ^2 were chosen to produce relationships between x and y accounting for 20%, 50% of the variation in y representing a range from fair to good fit of the line. Values of c were chosen so that the mean of the largest possible residuals or $f = c(x_{max} - x_{min})^2$ would take on values of 0, σ, 2σ, 3σ, 4σ, 5σ, 6σ, 7σ, 8σ. The obtained variances for $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$, and $\hat{c}_{ML}$ are given in Tables 1-3.

The asymptotic formula for the variance of $\hat{\alpha}_{ML}$ can be written as σ^2 times a function of γ, f, n and the spacing of the x's; it does not depend on α, β, or $\Delta = x_{max} - x_{min}$, the scaling of the x's. Therefore in Table 1 in which x's were equally spaced for all calculations, we show

$\frac{\text{var } \hat{\alpha}_{ML}}{\sigma^2}$ as $K(\gamma, f, n)$.

Examination of Table 1 shows that the asymptotic variance of $\hat{\alpha}_{ML}$ decreases monotonically from a maximum value at $c = 0$ but remains relatively stable across a wide range of f values from 2σ to 8σ. The variance of $\hat{\alpha}_{ML}$ increases with γ and decreases with n.

The asymptotic variance of $\hat{\beta}_{ML}$ can be written as $\frac{\sigma^2}{\Delta^2}$ times a function of γ, f, n and the spacing of the x's; it does not depend on α or β. Table 2 shows that the change in the variance of $\hat{\beta}_{ML}$ with f, γ, n is very similar to that for variance $\hat{\alpha}_{ML}$.

The asymptotic variance of $\hat{c}_{ML}$ can be written as $\frac{\sigma^2}{\Delta^4}$ times a function of γ,f, n and the spacing of the x's. The asymptotic variable of $\hat{c}_{ML}$ decreases rapidly as f increases until about 4σ or

5σ at which it begins to approach an asymptote. The variance of $\hat{c}_{ML}$ decreases as n increases but it also decreases as γ increases. For larger γ, the effective sample size for the estimation of c increases.

Table 1

Asymptotic Variance Formula for $\hat{\alpha}_{ML}$ Evaluated for Equally Spaced x's

$$\frac{\text{var}\ \hat{\alpha}_{ML}}{\sigma^2}$$

$\gamma = .20$ $\qquad c(x_{max} - x_{min})^2$

n	0	σ	2σ	3σ	4σ	5σ	6σ	7σ	8σ
21	.8798	.0963	.0659	.0626	.0620	.0613	.0607	.0602	.0599

$c(x_{max} - x_{min})^2 = 6\sigma$

n \ γ	.10	.20	.30	.40
15	.0760	.0847	.0946	.1067
21	.0544	.0607	.0679	.0767
41	.0279	.0312	.0350	.0396

Table 2

Asymptotic Variance of $\hat{\beta}_{ML}$ for Equally Spaced x's

$$\frac{\text{var } \hat{\beta}_{ML}}{\sigma^2}$$

$\gamma = .20$ $c(x_{max} - x_{min})^2$

n	0	σ	2σ	3σ	4σ	5σ	6σ	7σ	8σ
21	1.912	.2346	.1738	.1684	.1623	.1552	.1505	.1479	.1466

$c(x_{max} - x_{min})^2 = 6\sigma$

n \ γ	.10	.20	.30	.40
15	.1886	.2024	.2194	.2414
21	.1402	.1505	.1631	.1795
41	.0754	.0809	.0878	.0966

Table 3

Asymptotic Variance of $\hat{c}_{ML}$ for Equally Spaced x's

$$\frac{\text{var } \hat{c}_{ML}}{\sigma^2}$$

$\gamma = .20$ $c(x_{max} - x_{min})^2$

n	0	σ	2σ	3σ	4σ	5σ	6σ	7σ	8σ
21	11.14	.6545	.2249	.1490	.1204	.1055	.0974	.0931	.0907

$c(x_{max} - x_{min})^2$

n \ γ	.10	.20	.30	.40
15	.2502	.1315	.0963	.0823
21	.1860	.0974	.0713	.0609
41	.1000	.0522	.0381	.0325

SMALL SAMPLE PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATORS

We undertook a Monte Carlo study to investigate the properties of the maximum likelihood estimators of α, β, c in small samples. We set $\alpha = 0.0$, $\beta = 1.0$ and $m = x_{min}$ throughout. Eight parameter sets specifying the values of n, γ_T, σ^2, c and the spacing of the x's were defined and used to generate y samples, see Table 4. For each parameter set, evaluation of the properties of the estimators were made for several choices of γ_E, the value of γ actually used in estimation. The basic parameter set involved 21 equally spaced x's from 1 to 21, $\gamma_T = .20$, $\sigma^2 = 36$, and $c(x_{max} - x_{min})^2 = 6\sigma$. We chose $\sigma^2 = 36$ to obtain a representative situation in which x predicts 50% of the variance in y. The values $f = 6\sigma$ and $\gamma_T = .20$ were chosen because unless outliers are occasionally obvious by inspection it is unlikely that an outlier model would be applied (this is also approximately the value observed in the RJ data). The variations from this basic set of parameters include cases in which σ^2 is reduced, c is reduced, n is reduced, the x's follow a normal distribution, n is increased, and γ_T is varied.

For each parameter set and choice of γ_E, 200 random samples of y observations were generated using a random normal generator developed for the IBM 360 by Chen (1969). For some parameters, several sets of 200 samples were generated. The maximum likelihood estimators were obtained for each sample and the observed means and variances of the estimators across the 200 samples were calculated.

Table 4

Parameter Sets

$\alpha = 0.0 \quad \beta = 1.0 \quad m = x_{min}$

Set	n	γ	x's	σ^2	$c(x_{max}-x_{min})^2 c$	γ_E	No. of samples γ_E .01	.05	.10	.20	.30	.40
1. Basic	21	.20	equally spaced 1 to 21	36	6σ	.09		200	200	600	200	200
2. Reduce σ^2	21	.20	equally spaced 1 to 21	9	6σ	.045		200	200	400	200	200
3. Reduce c	21	.20	equally spaced 1 to 21	36	3σ	.045		200	200	200	200	200
4. Reduce n	15	.20	equally spaced 1 to 15	18.67	6σ	.1322		200	200	400	200	200
5. Vary x's	15	.20	expected normal order statistics	.85	6σ	.45		200	200	400	200	200
6. Increase n	41	.20	equally spaced 1 to 41	140	6σ	.04437		200	200	400	200	200
7. Reduce γ	21	.05	equally spaced 1 to 21	36	6σ	.09	200	200	200	200		
8. Increase γ	21	.40	equally spaced 1 to 21	36	6σ	.09			200	200	200	200

The initial estimates used in the iterative maximum likelihood solutions were $\hat{\alpha}_{LS}$, $\hat{\beta}_{LS}$, σ^2_{LS} , and c was estimated from the largest residual from the least squares line in the appropriate quadrant. In general the iterative solution converged to six significant digits in each estimator fairly rapidly. The procedure was automatically terminated after 100 iterations. Table 5 shows the number of iterations required for convergence for the basic parameter set with $\gamma_E = \gamma_T = .20$ and with $\gamma_E = .40$, $\gamma_E = .05$. The median number of iterations was in the range 20-29. The number of iterations required seems to increase somewhat as γ_E increases.

Tables 6, 7, and 8 show the results for $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$, and $\hat{c}_{ML}$, respectively. Part (a) of each table shows the ratio of the asymptotic variance to σ^2 for each parameter set for several γ values. (Note that this ratio is independent of σ^2). These figures have been scaled to allow comparisons with figures in Tables 1, 2, and 3 (i.e., they all correspond to calculations made for x ranging from -1 to +1.). Part (b) of each table shows the ratio of the observed variance using γ_E to the asymptotic variance calculated with γ_E . Part (c) of each table shows the ratio of the observed variance using γ_E to the asymptotic variance calculated with γ_T . Part (d) of each table shows the observed bias (due to scale changes these figures are not necessarily comparable from row to row). Part (e) shows the ratio of the squared bias to the asymptotic variance calculated with γ_T .

A guideline to the interpretation of the ratios between observed and asymptotic variances can be obtained by the following argument. If an estimator $\hat{\theta}$ is normally distributed, the standard deviation of its estimated variance based on p samples is $\sqrt{2/p}$ var $\hat{\theta}$. Thus the

Table 5

Number of Iterations Required for Convergence to Six Significant Digits in All Estimators for Basic Parameter Set

$\gamma_T = .20$

No. of Iterations	$\gamma_E = .20$ Frequency	$\gamma_E = .40$ Frequency	$\gamma_E = .05$ Frequency
1 - 9	1	0	2
10 - 19	63	25	94
20 - 29	82	69	53
30 - 39	27	33	21
40 - 49	7	22	14
50 - 59	5	13	9
60 - 69	2	7	1
70 - 79	4	8	1
80 - 89	0	3	4
90 - 99	2	4	0
100+	7	16	1
	200	200	200

standard deviation of $\frac{\widehat{\text{var } \hat{\theta}}}{\text{var } \hat{\theta}}$ is approximately $\sqrt{2/p}$. So for $p = 200$ we would expect the observed variances to be within $\pm 20\%$ of the true variance; for $p = 400$ and $p = 600$, the observed variances should be within $\pm 14\%$ or $\pm 12\%$ respectively.

Behavior when $\gamma_E = \gamma_T$

The properties of $\hat{\alpha}_{ML}$ are shown in Table 6. Note that the asymptotic variance of $\hat{\alpha}_{ML}$ is not strongly affected by c, γ, or the spacing of the x's. For the parameter sets investigated here the actual variance is only 14% to 35% larger than the asymptotic variance when $\gamma_E = \gamma_T$. The bias is generally positive but contributes less than 1% to the MSE $\hat{\alpha}_{ML}$.

The asymptotic variance of $\hat{\beta}_{ML}$ depends more heavily on c and the spacing of the x's than does the variance of $\hat{\alpha}_{ML}$. With the exception of two cases the observed variance is no more than 15% larger than the asymptotic variance. The bias is numerically quite small and makes a negligible contribution to MSE.

The asymptotic variance of $\hat{c}_{ML}$ is fairly strongly affected by changes in the parameters, especially by changes in γ . The observed variance is considerably larger than the asymptotic variance--about 2 to 6 times larger for these cases. The bias is generally negative indicating that c is underestimated on the average. The contribution of bias to MSE ranges from 4 to 22% except for the case where $\gamma_T = .05$.

How nearly normal are the distributions of $\hat{\alpha}$, $\hat{\beta}$, $\hat{c}$ in small samples? Histograms of the distributions of $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$, $\hat{c}_{ML}$ for the 600 samples generated by basic parameter set with $\gamma = .20$ $n = 21$ are shown in

Table 6

Properties of $\hat{\alpha}_{ML}$

a) Ratio of asymptotic variance of $\hat{\alpha}_{ML}$ to σ^2

Parameter set			γ					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2	.0485	.0513	.0544	.0607	.0679	.0767
2	21	.2	.0485	.0513	.0544	.0607	.0679	.0767
3	21	.2	.0486	.0515	.0550	.0626	.0716	.0828
4	15	.2	.0679	.0717	.0760	.0847	.0946	.1067
5	15	.2	.0681	.0726	.0777	.0882	.0998	.1137
6	41	.2	.0248	.0263	.0279	.0312	.0350	.0396
7	21	.05	.0485	.0513	.0544	.0607	.0679	.0767
8	21	.4	.0485	.0513	.0544	.0607	.0679	.0767

b) Ratio of observed variance using γ_E to asymptotic variance with γ_E

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		1.37	1.62	1.21	1.06	1.07
2	21	.2		2.02	1.38	1.14	1.24	.93
3	21	.2		1.08	1.41	1.27	1.28	.95
4	15	.2		1.46	1.66	1.34	1.08	1.29
5	15	.2		1.66	1.56	1.17	1.14	1.06
6	41	.2		1.63	1.06	1.18	.92	1.18
7	21	.05	1.12	1.19	1.03	.89		
8	21	.4			3.86	2.07	1.38	1.31

c) Ratio of observed variance using γ_E in estimation to asymptotic variance with γ_T

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		1.16	1.45	1.21	1.19	1.35
2	21	.2		1.72	1.23	1.14	1.39	1.17
3	21	.2		.89	1.24	1.27	1.47	1.24
4	15	.2		1.23	1.49	1.34	1.21	1.63
5	15	.2		1.37	1.38	1.17	1.29	1.37
6	41	.2		1.37	.95	1.18	1.03	1.52
7	21	.05	1.05	1.19	1.09	1.06		
8	21	.40			2.73	1.64	1.22	1.31

Table 6 (Continued)

Properties of $\hat{\alpha}_{ML}$

d) Bias in $\hat{\alpha}_{ML}$ when γ_E is used in estimation

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.20		.548	.532	.066	-.281	-.540
2	21	.20		.477	.268	.021	-.166	-.157
3	21	.20		.749	.200	.135	-.415	-.625
4	15	.20		.493	.365	.033	-.330	-.212
5	15	.20		.117	.099	.007	-.044	-.103
6	41	.20		1.354	.691	.072	-.595	-.947
7	21	.05	.024	-.047	.012	-.338		
8	21	.40			2.111	.931	.315	.056

e) Squared bias in $\hat{\alpha}_{ML}$ as a percent of the asymptotic variance using γ_T

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		14	13	0	4	14
2	21	.2		50	13	0	5	5
3	21	.2		25	2	1	8	17
4	15	.2		16	8	0	6	3
5	15	.2		18	14	0	3	14
6	41	.2		42	11	0	8	21
7	21	.05	0	0	0	6		
8	21	.40			162	31	4	1

Table 7

Properties of $\hat{\beta}_{ML}$

a) Ratio of asymptotic variance of $\hat{\beta}_{ML}$ to σ^2

Parameter set			γ					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2	.1314	.1354	.1402	.1505	.1631	.1795
2	21	.2	.1314	.1354	.1402	.1505	.1631	.1795
3	21	.2	.1327	.1409	.1500	.1684	.1893	.2148
4	15	.2	.1770	.1823	.1886	.2024	.2194	.2414
5	15	.2	.2389	.2468	.2559	.2752	.2983	.3276
6	41	.2	.0705	.0728	.0754	.0809	.0878	.0966
7	21	.05	.1314	.1354	.1402	.1505	.1631	.1795
8	21	.4	.1314	.1354	.1402	.1505	.1631	.1795

b) Ratio of observed variance of $\hat{\beta}_{ML}$ using γ_E to asymptotic variance using γ_E

Parameter set			γ					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		1.60	1.34	.97	1.11	1.20
2	21	.2		2.34	1.52	1.05	1.31	.88
3	21	.2		1.24	1.36	1.10	1.30	.89
4	15	.2		2.04	1.88	1.49	1.11	1.17
5	15	.2		1.95	1.56	1.14	1.18	1.34
6	41	.2		1.36	1.37	1.09	.81	.87
7	21	.05	1.14	1.15	.89	1.16		
8	21	.4			4.52	1.88	1.45	1.43

c) Ratio of observed variance when γ_E used in estimation to asymptotic variance using γ_T

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		1.44	1.25	.97	1.21	1.43
2	21	.2		2.12	1.42	1.05	1.43	1.05
3	21	.2		1.02	1.21	1.10	1.47	1.14
4	15	.2		1.83	1.74	1.49	1.20	1.39
5	15	.2		1.76	1.45	1.14	1.27	1.59
6	41	.2		1.22	1.28	1.09	.88	1.04
7	21	.05	1.10	1.15	.92	1.28		
8	21	.4			3.50	1.56	1.31	1.43

Table 7 (Continued)

Properties of $\hat{\beta}_{ML}$

d) Bias in $\hat{\beta}_{ML}$ when γ_E used in estimation

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.20		.0445	.0151	-.0007	.0095	-.0363
2	21	.20		.0444	.0160	.0075	.0108	.0125
3	21	.20		.0608	.0348	.0335	-.0540	.0560
4	15	.20		.0434	.0631	.0332	-.0189	-.0473
5	15	.20		.0507	.0187	.0038	-.0091	-.0414
6	41	.20		.0226	.0308	-.0035	-.0061	-.0171
7	21	.05	.0009	.0236	-.0266	-.0027		
8	21	.40			.2201	.0716	.0283	.0186

e) Squared bias in $\hat{\beta}_{ML}$ as a percent of the asymptotic variance using γ_T

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		4	0	0	0	0
2	21	.2		15	2	1	1	0
3	21	.2		5	2	2	5	3
4	15	.2		3	6	2	0	3
5	15	.2		4	0	0	0	2
6	41	.2		2	3	0	0	1
7	21	.05	0	1	1	0		
8	21	.4			74	8	2	1

Table 8

Properties of $\hat{c}_{ML}$

a) Ratio of asymptotic variance of $\hat{c}_{ML}$ to σ^2

Parameter set			γ					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2	2.0367	.3754	.1860	.0974	.0713	.0609
2	21	.2	2.0367	.3754	.1860	.0974	.0713	.0609
3	21	.2	4.0500	.6515	.3046	.1490	.1040	.0859
4	15	.2	2.7203	.5037	.2502	.1315	.0963	.0823
5	15	.2	3.6619	.6705	.3309	.1721	.1252	.1065
6	41	.2	1.1051	.2024	.1000	.0522	.0381	.0325
7	21	.05	2.0367	.3754	.1860	.0974	.0713	.0609
8	21	.4	2.0367	.3754	.1860	.0974	.0713	.0609

b) Ratio of observed variance using γ_E to asymptotic variance using γ_E

Parameter set			γ					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		1.07	2.90	4.47	5.30	8.55
2	21	.2		1.29	2.46	5.25	6.49	6.72
3	21	.2		.55	1.04	1.95	6.07	2.74
4	15	.2		1.46	1.89	6.10	5.30	7.58
5	15	.2		1.29	2.36	3.78	5.08	4.80
6	41	.2		.99	1.50	2.97	5.95	8.41
7	21	.05	.55	2.77	6.12	9.90		
8	21	.4			1.24	1.19	1.37	3.32

c) Ratio of observed variance when γ_E used in estimation to asymptotic variance using γ_T

Parameter set			γ					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		4.50	5.51	4.47	3.86	5.33
2	21	.2		4.97	4.69	5.25	4.74	4.20
3	21	.2		2.41	2.31	1.95	4.24	1.58
4	15	.2		5.57	3.61	6.10	3.88	4.74
5	15	.2		5.04	4.54	3.78	3.70	2.97
6	41	.2		3.84	2.87	2.97	4.35	5.24
7	21	.05	3.00	2.77	3.03	2.56		
8	21	.4			3.79	1.90	1.60	3.32

Table 8 (Continued)

Properties of $\hat{c}_{ML}$

d) Bias in $\hat{c}_{ML}$ when γ_E used in estimation

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.20		-.0071	-.0131	-.0087	-.0116	-.0149
2	21	.20		-.0025	-.0026	-.0028	-.0081	-.0087
3	21	.20		-.0072	-.0018	-.0071	.0050	-.0082
4	15	.20		-.0121	-.0217	-.0138	-.0158	-.0295
5	15	.20		-.0525	-.0493	-.0570	-.0397	-.0832
6	41	.20		-.0017	.0015	-.0018	-.0042	-.0067
7	21	.05	-.0347	-.0364	-.0363	-.0339		
8	21	.40			-.0072	-.0021	-.0003	-.0032

e) Squared bias in $\hat{c}_{ML}$ as a percent of the asymptotic variance using γ_T

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		14	49	22	38	63
2	21	.2		7	8	9	75	87
3	21	.2		10	1	9	4	13
4	15	.2		15	46	19	24	86
5	15	.2		17	15	20	10	43
6	41	.2		6	5	7	39	99
7	21	.05	88	98	98	75		
8	21	.4			24	2	0	4

Figures 2-4. These distributions appear reasonably symmetric and well-behaved, especially the distribution of $\hat{\beta}_{ML}$. Note the second peak at zero in the distribution of $\hat{c}$.

In summary then, for γ known, $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ appear to behave well for samples as small as 21. The bias is not large and the asymptotic variance formula if inflated by 20 to 50% could reasonably be used to provide some estimates of precision. The estimator of c performs poorly by contrast, it is an underestimate on the average and much more variable than asymptotic results would indicate. This is hardly surprising since the effective sample size for the estimation of c is of the order of $\frac{\gamma n}{2}$.

Robustness to Inaccurate Specification of γ

Now that we have assessed the small sample behavior of $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$, and $\hat{c}_{ML}$ when the true γ is known we need to evaluate how misled we will be if the wrong value of γ is used in the estimation procedure. For each parameter set we have run sets of 200 samples when the value of γ used in the estimation procedure, γ_E , is not equal to the true γ value, γ_T . Changes in the observed variance and observed bias due to inaccurate specification of γ are shown in Tables 6, 7, 8.

For $\hat{\alpha}_{ML}$ we note that inaccurate specification of γ does not seem to have an appreciable effect on the size of the variance. Table 7t shows that the ratio of the observed to the asymptotic variance using γ_E was generally less than 1.7. The bias in $\hat{\alpha}_{ML}$ is much more directly affected by γ_E ; it is quite close to zero when $\gamma_E = \gamma_T$, becoming moderate and positive for $\gamma_E < \gamma_T$ (γ underestimated), and moderate and negative for $\gamma_E > \gamma_T$ (γ overestimated). That is, underestimation of

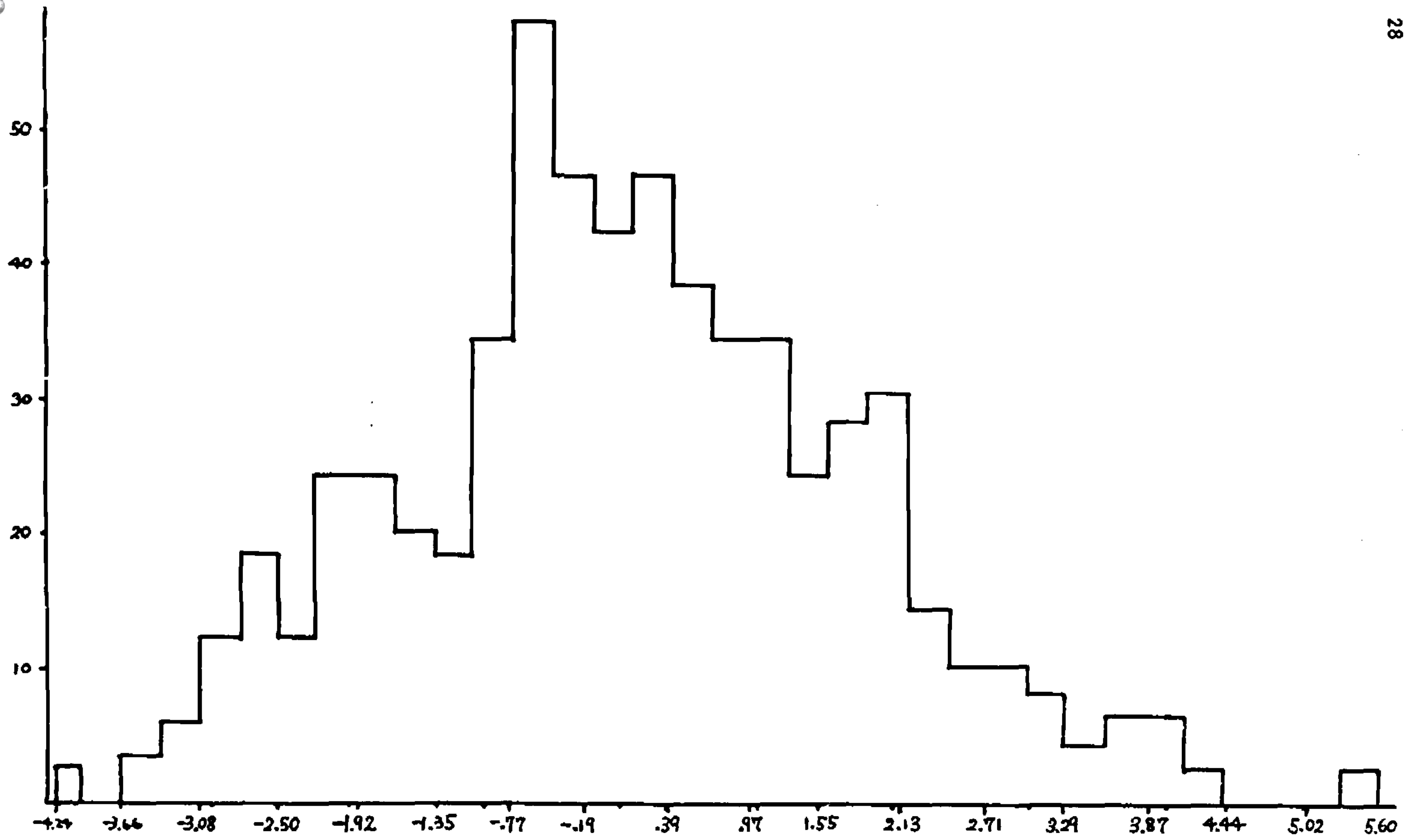

Figure 2: Distribution of $\hat{\alpha}_{ML}$ in 600 samples using the basic parameter set with $\gamma_E = \gamma_T = .20$

Observed
mean = .066
variance = 2.64
skewness = .185
kurtosis = -.057

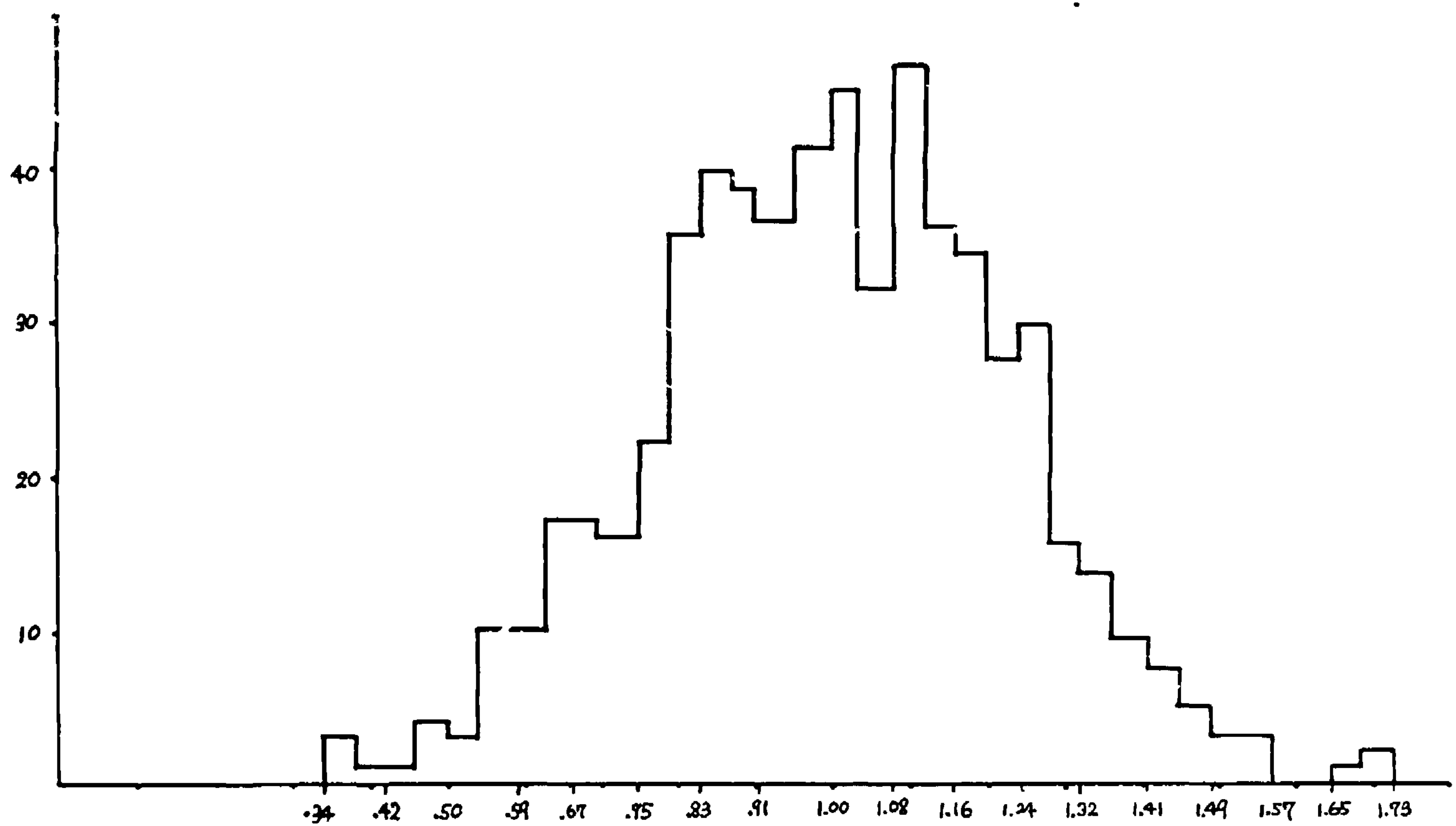

Figure 3: Distribution of $\hat{\beta}_{ML}$ in 600 samples using the basic parameter set with $\gamma_E = \gamma_T = .20$

Observed
mean = .999
variance = .0526
skewness = -.004
kurtosis = -.04

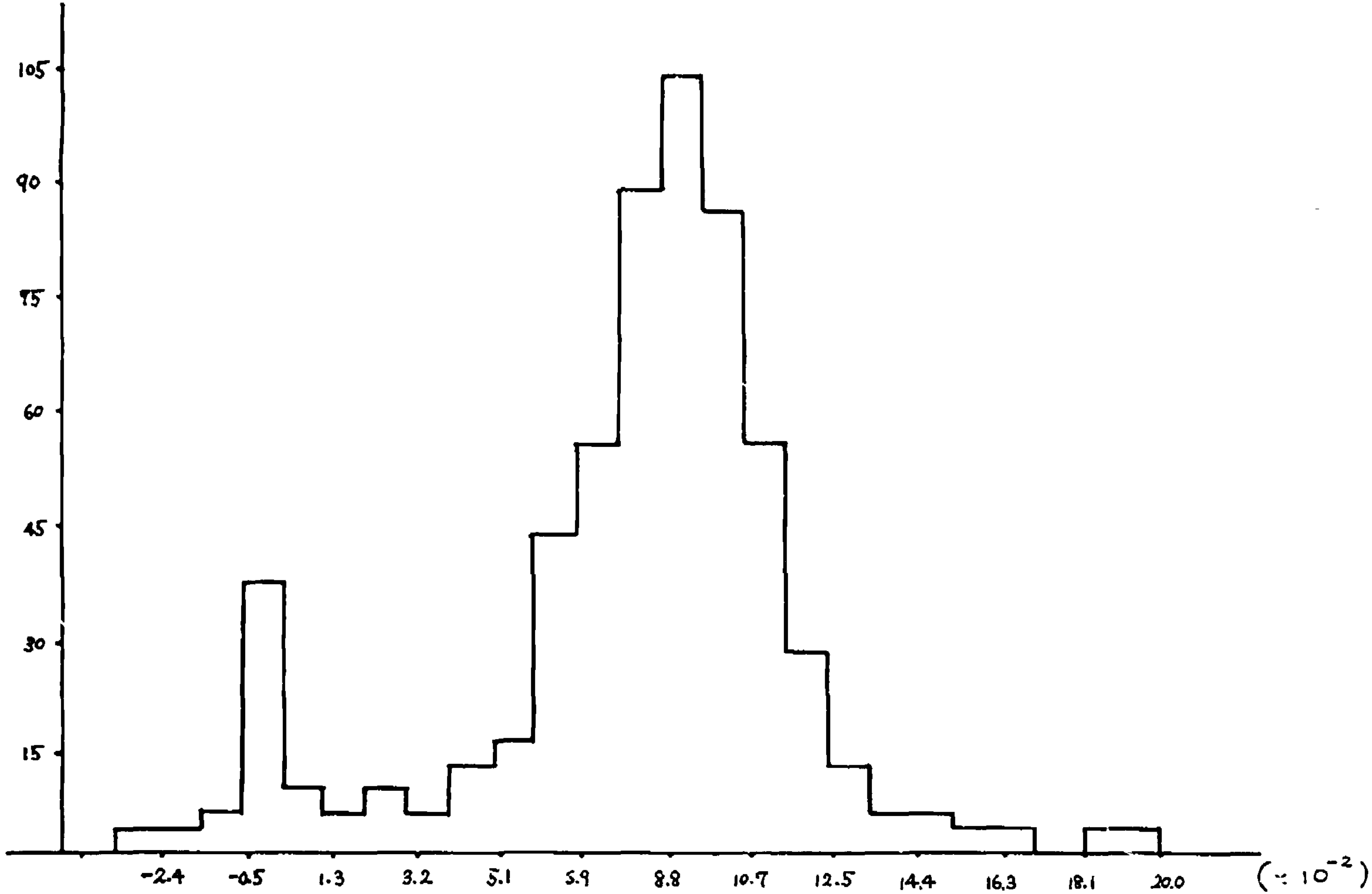

Figure 4: Distribution of $\hat{c}_{ML}$ in 600 samples using the basic parameter set with $\gamma_E = \gamma_T = .20$

Observed
mean = .081
variance = .00157
skewness = -.56
kurtosis = 1.59

γ leads to overestimation of α and overestimation of γ leads to underestimation of α. Although using a γ of .05 or .40 when the true value is .20 is quite a large error, the bias at these extremes generally contributes only about 20% to the MSE.

The situation for $\hat{\beta}_{ML}$ is very similar although inaccurate specification of γ seems to have a somewhat larger effect on the variance. Again, the bias tends to switch from positive to negative as we go from underestimation to overestimation of γ; however, the bias is generally of negligible size even at the extremes.

For $\hat{c}_{ML}$, inaccurate specification of γ does not exhibit any appreciable tendency to inflate the variance. The observed variance is much more strongly influenced by the value of γ_T than by γ_E, tending to be comparatively stable across a row. The effect on the size of the consistently negative bias is variable, with overestimation of γ considerably worse than underestimation.

In summary then, even for relatively small samples the maximum likelihood estimators of α, β, c are robust to inaccurate specification of γ. Their variances are only moderately affected by differences between γ_E and γ_T, and bias becomes a serious problem only for $\hat{c}_{ML}$ when γ is overestimated.

COMPARISON OF MAXIMUM LIKELIHOOD ESTIMATORS AND ORDINARY LEAST SQUARES ESTIMATORS OF α, β

How much can we generally expect to gain by using the maximum likelihood estimators of $\hat{\alpha}$, $\hat{\beta}$ rather than the ordinary least squares estimators whose computation ignores the presence of outliers. The ordinary least squares estimators are given by

$$\hat{\alpha}_{ls} = \bar{y}$$

$$\hat{\beta}_{ls} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$\sigma^2_{ls} = \frac{\Sigma(y_i - \bar{y} - \beta_{ls}\ (x_i - \bar{x}))^2}{n-2}$$

To derive the expected values and variances of these estimators under the quadratic outlier model we note that

$$E(y_i) = \alpha + \beta(x_i - \bar{x}) + \gamma c\ (x_i - m)^2$$

and

$$\text{Var}\ (y_i) = \sigma^2 + \gamma(1-\gamma)c^2\ (x_i - m)^4\ .$$

Thus under the outlier model

$$E(\hat{\alpha}_{ls}) = \alpha + \frac{\gamma c}{n}\,\Sigma(x_i - m)^2$$

$$\text{Var}\ (\hat{\alpha}_{ls}) = \frac{\sigma^2}{n} + c^2\,\frac{\gamma(1-\gamma)}{n^2}\,\Sigma(x_i - m)^4$$

$$E(\hat{\beta}_{1s}) = \beta + \gamma c \frac{\Sigma(x_i - \bar{x})(x_i - m)^2}{\Sigma(x_i - \bar{x})^2}$$

$$\text{Var } (\hat{\beta}_{1s}) = \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2} + c^2 \gamma(1 - \gamma) \frac{\Sigma(x_i - \bar{x})^2 (x_i - m)^4}{[\Sigma(x_i - \bar{x})^2]^2}$$

$$E(\hat{\sigma}^2_{1s}) = \sigma^2 + \frac{\gamma c^2}{n-2} \Sigma(x_i - m)^4 - \frac{\gamma c^2}{n(n-2)} [(1-\gamma)\Sigma(x_i - m)^4 + \gamma[\Sigma(x_i - m)^2]^2]$$

$$- \frac{\gamma c^2}{\Sigma(x_i - \bar{x})^2(n-2)} [(1-\gamma)\Sigma(x_i - \bar{x})^2(x_i - m)^4 + \gamma[\Sigma(x_i - \bar{x})(x_i - m)^2]^2] .$$

The estimators of α and β are inflated by terms in γc and the x's, their variances are increased by terms in $c^2\gamma(1-\gamma)$ and the x's.

The maximum improvement obtainable from using the maximum likelihood estimators can be assessed by looking at $\frac{\text{MSE } \hat{\alpha}_{1s}}{\text{MSE } \hat{\alpha}_{ML}}$ and $\frac{\text{MSE } \hat{\beta}_{1s}}{\text{MSE } \hat{\beta}_{ML}}$ where the asymptotic formulas for the ML estimators are used. (Since $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$ are asymptotically unbiased, MSE = var .) Calculations for $\alpha = 0$, $\beta = 1$, equally spaced x's between -1.0 and $+1.0$, $m = -1.0$, are displayed for several values of γ, n, c in Table 9. Improvement from using the ML estimators is rapid with increases in γ, c, n. For $n = 21$, $\gamma = .2$, $f = 6\sigma$, the mean squared error using the least squares estimators is almost five times that using the maximum likelihood estimators.

The ratios of mean squared errors observed in the Monte Carlo study are shown in Tables 10 and 11. Although the observed advantage of the maximum likelihood estimators of α and β is less than indicated

Table 9

a) Asymptotic formulas for $\dfrac{\text{MSE}\ (\hat{\alpha}_{LS})}{\text{MSE}\ (\hat{\alpha}_{ML})}$

$\gamma = .20$

$c(x_{max} - x_{min})^2$

	0	σ	2σ	3σ	4σ	5σ	6σ
n = 21	.06	.56	1.11	1.67	2.39	3.34	4.52

$c(x_{max} - x_{min})^2 = 6\sigma$

n	γ	.1	.2
21		2.26	4.52
41		2.90	6.97

b) Asymptotic formulas for $\dfrac{\text{MSE}\ (\hat{\beta}_{LS})}{\text{MSE}\ (\hat{\beta}_{ML})}$

$\gamma = .20$

$c(x_{max} - x_{min})^2$

	0	σ	2σ	3σ	4σ	5σ	6σ
n = 21	.07	.62	1.14	1.63	2.47	3.59	4.96

$c(x_{max} - x_{min})^2 = 6\sigma$

n	γ	.1	.2
21		2.57	4.96
41		3.08	6.95

Table 10

Observed $\frac{\text{MSE } \hat{\alpha}_{LS}}{\text{MSE } \hat{\alpha}_{ML}}$

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		3.09	3.34	3.90	3.45	2.67
2	21	.2		2.34	3.68	4.01	2.70	3.96
3	21	.2		1.39	1.37	1.36	1.08	1.04
4	15	.2		2.47	2.52	3.05	2.64	2.30
5	15	.2		1.87	2.24	2.48	2.61	2.23
6	41	.2		3.52	6.78	8.31	6.06	4.01
7	21	.05	1.24	1.22	1.44	1.20		
8	21	.4			2.68	5.67	8.72	7.95

Table 11

Observed $\frac{\text{MSE } \hat{\beta}_{LS}}{\text{MSE } \hat{\beta}_{ML}}$

Parameter set			γ_E					
No.	n	γ_T	.01	.05	.10	.20	.30	.40
1	21	.2		3.02	4.09	4.78	4.28	3.36
2	21	.2		2.55	3.87	4.76	3.47	4.94
3	21	.2		1.28	1.36	1.68	1.12	1.39
4	15	.2		2.37	2.77	3.44	3.53	2.77
5	15	.2		2.20	2.35	2.92	2.69	2.74
6	41	.2		4.36	5.92	6.33	8.06	6.28
7	21	.05	1.38	1.61	1.28	1.48		
8	21	.4			2.86	6.77	8.81	7.46

by asymptotic results, it is still considerable. The MSE using standard least squares is at least 2.4 that using the maximum likelihood estimators with the true γ for all but the cases with $f = 3\sigma$ and $\gamma = .05$. The maximum likelihood estimators still perform better than the least squares estimators even when the estimated γ is way off. The advantage of the maximum likelihood estimators increases rapidly with small increases in sample size.

Comparisons were also made between the maximum likelihood estimators and the ordinary least squares estimators for a quadratic regression. However, for x's symmetric about $\bar{x}$, the two least squares estimators are identical and MSE $\hat{\alpha}$ was little different in the two situations.

APPLICATIONS

I became interested in the problem of outliers in regression when I undertook with Professor Richard Snow at Stanford a reanalysis of the data reported on by Rosenthal and Jacobson in their book Pygmalion in the Classroom. All the children in a particular grade school were given a preliminary I.Q. test. Then, one-fifth of the children were selected at random and their teachers told that these experimental children were expected to bloom intellectually very soon. Months later all the children, both experimental and control groups, were retested with the same IQ test. One way to assess differences between the two groups is to compare the regression of posttest IQ on pretest IQ. However, we soon found that although a straight line seemed to describe the majority of children fairly well, some children had excessively high IQ's on the retest.

Look at Figure 5 which shows pre and post Total IQ scores for the 19 experimental group children in the first and second grades. One child with a pretest IQ score of 139 has a posttest IQ score of 202. Figure 6 shows the Verbal IQ results for the third- and fourth-grade experimental group. Figure 7 shows the Verbal IQ results for the fifth- and sixth-grade experimental group. Other similar patterns appear for other groups in the experiment; except for the first- and second-grade Reasoning subtest, which has some excessively low pretest scores, the general picture is the same for Verbal and Reasoning subtests for all grades. Most of the points seem to lie on a straight line, but some children with high pretest scores have excessively high posttest scores. Thus we have a problem where outliers seem to occur only for high values of x .

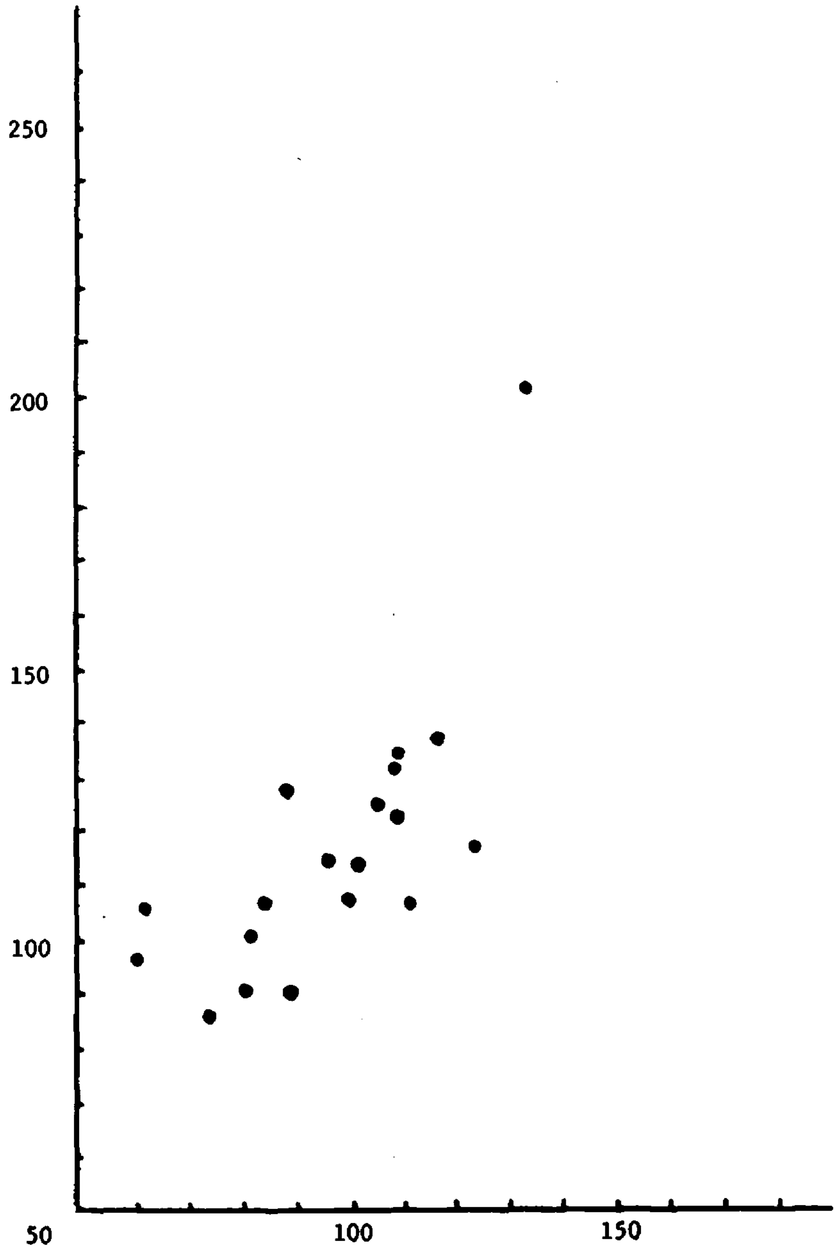

Figure 5: Pre and Post Total IQ scores for 19 experimental group children in grades 1 & 2 (Note that both scales start at 50.)

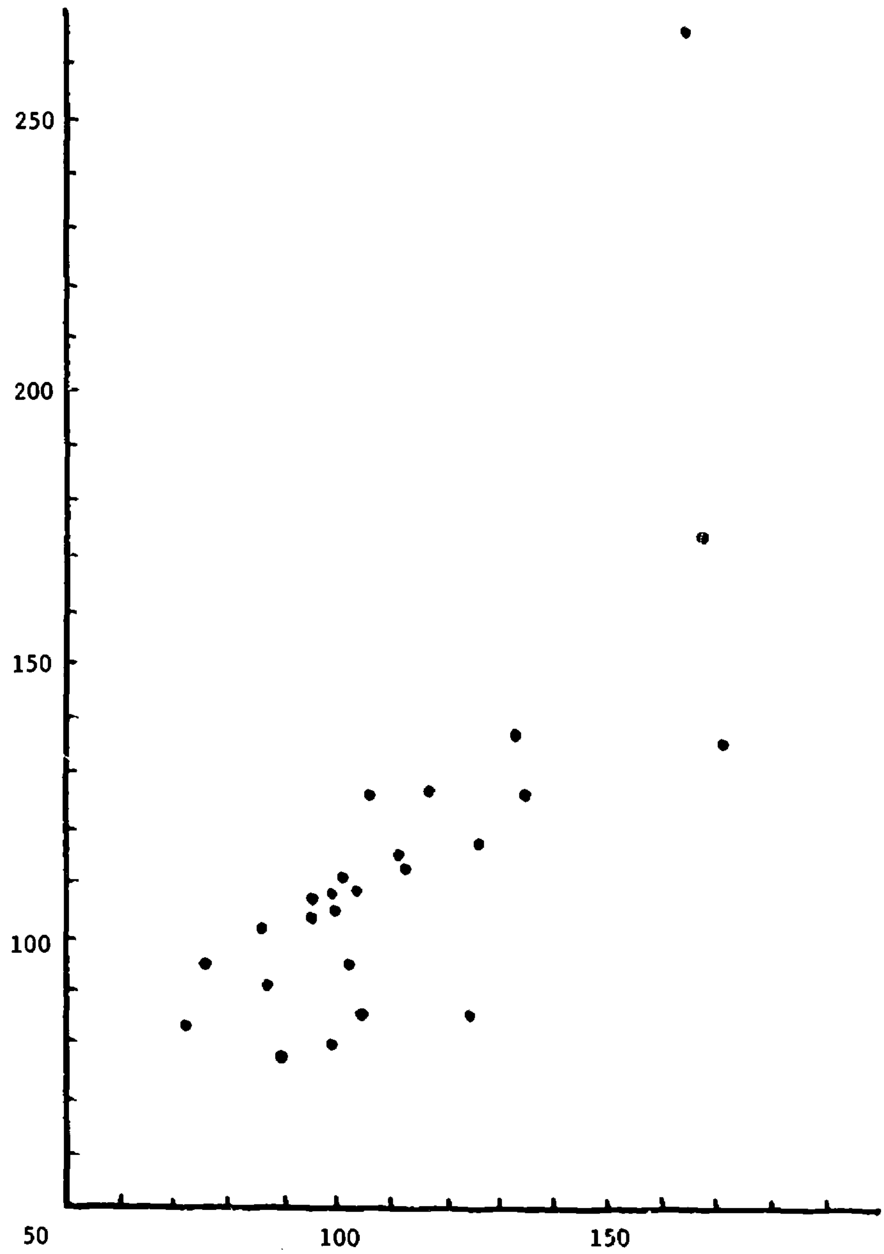

Figure 6: Pre and Post Verbal IQ scores for 26 experimental group children in grades 3 & 4 (Note that both scales start at 50.)

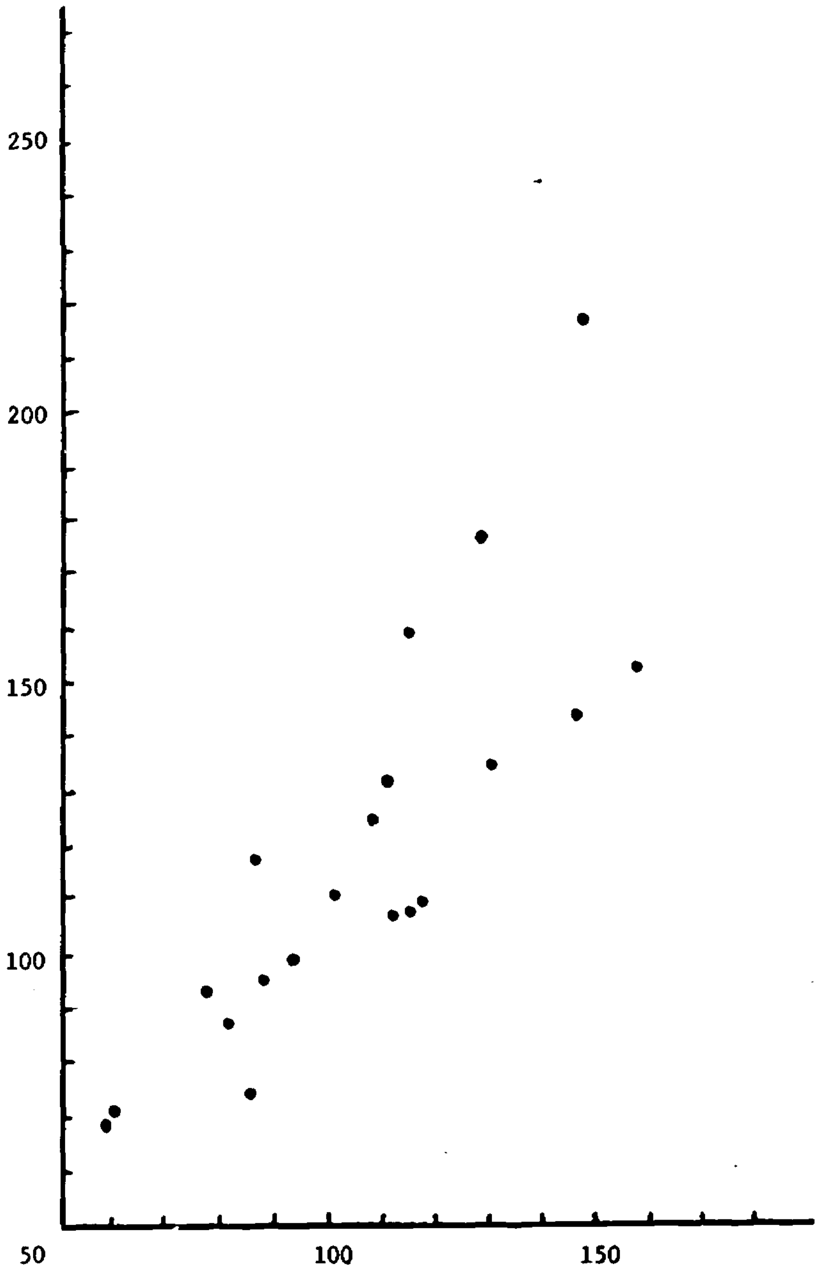

Figure 7: Pre and Post Verbal IQ scores for 23 experimental group children in grades 5 & 6 (Note that both scales start at 50.)

In this problem the apparent outliers may be partially due to the IQ transformation of the raw scores. At the extremes of the score distribution, one question right or wrong can make many points difference in IQ. However, the raw scores were no longer available to us, and it is the IQ scores which generally receive the psychological interpretation.

We then applied our outlier model to the estimation of α and β for three sets of data from the Rosenthal experiments. Tables 12, 13, and 14 show the results for grades 1 and 2, grades 3 and 4, and grades 5 and 6, the data shown in the scatterplots. The iterative solutions of the maximum likelihood equations converged to at least six significant digits after 20-25 iterations.

Look first at Table 12. When standard least squares was used we obtained $\hat{\alpha} = 117$, $\hat{\beta} = .93$ and $s^2 = 376$. When the one "obvious outlier" was removed $\hat{\alpha} = 112$, $\hat{\beta} = .58$ and $s^2 = 159$ using standard least squares. The maximum likelihood estimates obtained with $\gamma = .05$ are $\hat{\alpha} = 113$, $\hat{\beta} = .58$, $s^2 = 141$, $c = .0126$. Notice that these estimates change very little for values of γ ranging from .01 to .20, and how similar they are to those obtained by deleting the outlier and using standard least squares. The estimate of $\hat{c}$ is very close to that obtained by fitting the bias term through the outlier point. We also tried $\gamma = .001$ and obtained $\hat{\alpha} = 116.4$, $\hat{\beta} = .90$, $s^2 = 339$ -- very similar to standard least squares on all the data.

For grades 3 and 4 the data resemble the grades 1 and 2 data with one outlier, but there are two y points near the line for very large x; that is, the basic line appears better defined, and the outlier is farther out. Here even for $\gamma = .001$ the results were very little

affected by choice of γ and resembled those for standard least squares with the outlier deleted.

Our results for grades 5 and 6 were very similar. The choice of γ had little effect on the estimates of α, β, c for $\gamma = .01$ to $.30$; s^2 was most affected. For $\gamma = .001$ results were close to the standard least squares on all the data. These data do not look like a one-outlier problem and the results obtained using our method do not resemble those obtained by deleting one outlier. These data look much more like our second interpretation--a mixture of linear and quadratic regression.

Our estimates of c were similar for all three pieces of data: .0126, .0112, and .0102.

In conclusion the model seems general enough to represent many outlier problems. Choice of γ in any reasonable range seems to make little difference in the estimates. The data seem to dominate the specification of γ. Use of this model has reflected well our intuitive impressions of the data.

In general it seems desirable to fit a model which describes all the data well--outliers and all--and regression problems with outliers dependent on the x value could use considerable investigation.

Table 12

Regression Analyses for Grades 1 & 2

Experimental Group Total IQ 1 & 3, N = 19

Standard Least Squares

	$\hat{\alpha}$	$\hat{\beta}$	$s_{y \cdot x}$
All Data	116.7	.93	19.39
Outlier reduced from 202 to 160	114.5	.71	13.48
Outlier deleted	112.0	.58	12.63

Maximum Likelihood Estimates Under Outlier Model (m = 60)

γ	$\hat{\alpha}$	$\hat{\beta}$	s^2	$\hat{c}$
.001	116.4	.8997	338.96	.0098
.01	113.13	.5771	141.59	.0127
.05	112.92	.5785	141.49	.0126
.10	112.65	.5796	142.13	.0124
.20	111.97	.5782	146.13	.0120
Outlier deleted				
γ = .05	111.82	.5698	150.01	.0024

Solutions converged to 6 significant figures after about 20 iterations.

Table 13

Regression Analyses for Grades 3 & 4

Experimental Group Verbal IQ 1 & 3, N = 26

Standard Least Squares

	$\hat{\alpha}$	$\hat{\beta}$	$s_{y \cdot x}$
All Data	115.65	1.07	26.92
1 Outlier deleted	109.60	.70	13.85

Maximum Likelihood Estimates Under Outlier Model (m = 60)

γ	$\hat{\alpha}$	$\hat{\beta}$	s^2	$\hat{c}$
.001	11			
.01	110.97	.7052	191.52	.0113
.05	110.70	.7123	191.48	.0112
.10	110.35	.7214	191.66	.0112
.20	109.59	.7405	192.92	.0110
1 Outlier deleted				
.05	109.41	.6708	190.93	.0022

Table 14

Regression Analyses for Grades 5 & 6

Experimental Group Verbal IQ 1 & 3, N = 23

Standard Least Squares

	$\hat{\alpha}$	$\hat{\beta}$	$s_{y \cdot x}$	s^2
All Data	115.35	1.14	20.8	432.6
1 Outlier deleted	110.73	.98	15.9	252.8

Maximum Likelihood Estimates Under Outlier Model (m = 60)

γ	$\hat{\alpha}$	$\hat{\beta}$	s^2	$\hat{c}$
.001	115.24	1.137	408.7	.007
.01	108.65	.8455	97.40	.0102
.05	108.27	.8444	93.07	.0102
.10	107.83	.8433	88.14	.0103
.20	106.99	.84	81.14	.0103
.30	106.31	.8405	76.13	.0104
.60	105.0	.8468	71.66	.0105
.70	104.6	.8503	71.95	.0105
.80	104.3	.8544	73.25	.0105
1.0	114.4	1.136	414	
1 Outlier deleted				
.05	107.72	.8459	95.64	.0112

CONCLUSIONS

We have proposed a model describing outliers in a linear regression problem, derived the maximum likelihood estimators of the parameters, and examined the asymptotic properties of the estimators. We have examined the behavior of the estimators in small samples and their robustness to inaccurate specification of γ . We have applied the model to some real data.

Although this particular quadratic outlier model was suggested by some real data and seems to be useful in the analysis of that data, the importance of the paper does not lie in this particular model but in the demonstration that models of this kind can be useful, that the asymptotic properties provide not unreasonable indications of behavior for samples as small as 21, and that the procedure may be quite robust to inaccurate specification of γ . Thus, building models to describe outliers and estimating the parameters of these models provides an interesting alternative to procedures of outlier detection followed by ordinary least squares procedures.

References

Adichie, J. N. (1967a), "Estimates of Regression Parameters Based on Rank Tests," Ann. Math. Statist., 38, 894-904.

Adichie, J. N. (1967b), "Asymptotic Efficiency of a Class of Nonparametric Tests for Regression Parameters," Ann. Math. Statist., 38, 884-893.

Anscombe, F. J. (1960), "Rejection of Outliers," Technometrics, 2, 123-146.

Anscombe, F. J. (1967), "Topics in the Investigation of Linear Relations Fitted by the Method of Least Squares," J. Royal Statist. Soc., B, 29, 1-52.

Anscombe, F. J. and B. A. Barron (1966), "Treatment of Outliers in Samples of Size Three," Journal of Research of the National Bureau of Standards, B, 70, 141.

Barnett, Roy N. (1965), "A Scheme for the Comparison of Quantitative Methods," The American Journal of Clinical Pathology, 43, 562-569.

Box, G. E. P. and G. C. Tiao (1968), "A Bayesian Approach to Some Outlier Problems," Biometrika, 55, 119-130.

Chen, H. T. (1970), "A Random Normal Number Generator for 32-bit Word Computers," submitted to JASA.

Day, N. E. (1969), "Estimating the Components of a Mixture of Normal Distributions," Biometrika, 56, 463-474.

Dixon, W. J. (1962), "Rejection of Observations," in Contributions to Order Statistics (A. E. Sarhan and B. G. Greenberg, editors), 299-342. New York: Wiley.

Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regression," Ann. Math. Statist., 34, 447-456.

Ferguson, Thomas S. (1961a), "Rules for Rejection of Outliers," Revue Inst. Int. de Stat., 29:3, 29-43.

Ferguson, Thomas S. (1961b), "On the Rejection of Outliers," in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Jerzy Neyman, editor). Berkeley: U. C. Press.

Grubbs, Frank E. (1950), "Sample Criteria for Testing Outlying Observations," Ann. Math. Statist., 21, 27-58.

Grubbs, Frank E. (1969), "Procedures for Detecting Outlying Observations in Samples," Technometrics, 11, 1-22.

Hill, Bruce M. (1963), "Information for Estimating the Proportions in Mixtures of Exponential and Normal Distributions," JASA, 58, 918-932.

Karlin, Samuel and Donald Truax (1960), "Slippage Problems," Ann. Math. Statist., 31, 296-324.

Mickey, M. R., O. J. Dunn, and Virginia Clarke (1967), "Note on the Use of Stepwise Regression in Detecting Outliers," Computers and Biomedical Research, 1, 105-111.

Mood, A. M. (1950), Introduction to the Theory of Statistics. New York: McGraw Hill.

Pearson, K. P. (1894), "Contributions to the Mathematical Theory of Evolution. 1. Dissection of Frequency Curves," Philos. Trans. Roy. Soc., A, 185, 71.

Rao, C. R. (1952), Advanced Statistical Methods in Biometric Research. New York: Wiley.

Rosenthal, Robert and Lenore Jacobson (1968), Pygmalion in the Classroom. New York: Holt, Rinehart and Winston.

Sen, P. K. (1968), "Estimates of the Regression Coefficient Based on Kendall's Tau," JASA, 63, 1379-1389.

Theil, H. (1950), "A Rank Invariant Method of Linear and Polynomial Regression Analysis," Indag. Math., 12, Facs. 2, 85-91, 173-177.

Thomas, E. A. C. (1969), "Distribution Free Tests for Mixed Probability Distributions," Biometrika, 56, 475-484.

Wilks, S. (1963), "Multivariate Statistical Outliers," Sankhya, A, 25, 407-426.